6. International Committee on Taxonomy of Viruses and King A.M.Q. Academic Press; 2012. Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses.

7. Altschul S.F. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs / S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman // Nucleic Acids Res. 1997 - Sep 1;25(17):3389-402.

8. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792-7.

9. Tamura K. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 / K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar // Mol. Biol. Evol. 2013 - V. 30. P. 2725-2729.

DOI:10.34617/b0tr-5x23
УДК 574.24

ПОИСК ГОМОЛОГОВ S2-БЕЛКА ШИПИКОВ БЫЧЬЕГО КОРОНАВИРУСА В МЕТАГЕНОМАХ ОКЕАНА И СИЛОСА, И АНАЛИЗ ИХ ФИЛОГЕНИИ МЕТОДОМ UPGMA

Зимин Андрей Антонович[1], канд. биол. наук
Карманова Александра Николаевна[1,2]
Осепчук Денис Васильевич[3,4], д-р с.-х. наук[3]
[1]Институт биохимии и физиологии микроорганизмов им. Г.К. Скрябина РАН – обособленное подразделение ФИЦ «Пущинский научный центр биологических исследований РАН», г. Пущино, Российская Федерация
[2]Вятский государственный Российская университет, г. Киров
[3]ФГБНУ «Краснодарский научный центр по зоотехнии и ветеринарии», г. Краснодар, Российская Федерация
[4]ФГБОУ ВО «Кубанский государственный аграрный университет имени И.Т. Трубилина», г. Краснодар, Российская Федерация

В данной работе проведен поиск океанических гомологов белка S2 шипиков бычьего коронавируса. В метагеномах планктонической и осадочной микробиот мирового океана было найдено 15 и 4 гомологов этого вирусного белка соответственно. Выбор S2-гликопротеина в качестве объекта исследования базировался на нескольких факторах – схожести его структуры у коронавирусов разных родов и высокой консервативности их последовательностей. Анализ океанических гомологов поможет лучше понять пути происхождения исследуемого белка. Филогения данного набора аминокислотных последовательностей была исследована методом UPGMA.

Ключевые слова: BCoV; бычий коронавирус; SARS-COV-2; COVID-19; SARS-COV; MERS-COV; коронавирус; бетакоронавирус; коронавирусы; микробиота пелагики океана; микробиота наземных животных; метагеномика; метод UPGMA

SEARCH FOR HOMOLOGS OF S2-PROTEIN OF BOVIN CORONOVIRUS IN OCEAN METAGENOMES, AND ANALYSIS OF THEIR PHYLOGENY BY UPGMA METHOD

Zimin Andrei Antonovich[1], PhD Biol. Sci.
Karmanova Aleksandra Nikolaevna[1,2]
Osepchuk Denis Vasilyevich[3,4], Dr. Agr. Sci.
[1]Institute of Biochemistry and Physiology of Microorganisms named after G. K. Scriabin RAS – a separate subdivision of the Federal Research Center «Pushchino Scientific Centre for Biological Research of the Russian Academy of Sciences», Pushchino, Russian Federation
[2]Vyatka State University, Kirov, Russia
[3]Krasnodar Research Centre for Animal Husbandry and Veterinary Medicine, Krasnodar, Russian Federation
[4]Kuban State Agrarian University named after I. T. Trubilin, Krasnodar, Russian Federation

In this work, we search for oceanic homologues of the S2 protein of the bovine coronavirus spines. In the metagenomes of the planktonic and sedimentary microbiota of the world ocean, 15 and 4 homologs of this viral protein, respectively, were found. The choice of S2-glycoprotein as an object of study is based on several factors – the similarity of the coronaviruses structures of different genera, high conservatism of their sequences. An analysis of oceanic homologs will help to better understand the origin of the protein under study. The phylogeny of this set of amino acid sequences was studied using the UPGMA method.

Key words: BCoV; bovine coronavirus; SARS-COV-2; COVID-19; SARS-COV; MERS-COV; coronavirus; betacoronavirus; coronaviruses; ocean pelagic microbiota; terrestrial animal microbiota; metagenomics; UPGMA method

Первый случай нового коронавируса был зарегистрирован 30 декабря 2019 года в городе Ухань, провинция Хубэй, П.Р. Китай. Быстрые действия были предприняты Центром по контролю и профилактике заболеваний (CDC), китайскими органами здравоохранения и исследователями. Всемирная организация здравоохранения (ВОЗ) временно назвала этот патоген новым коронавирусом 2019 года (2019-nCoV) . 12 февраля ВОЗ окончательно назвала патоген 2019-nCoV как SARS-CoV-2, а вызывающее заболевание – как болезнь коронавируса 2019 (COVID-2019). Быстрые действия китайского правительства помогли им контролировать COVID-19 в Китае [1]. Тем не менее, SARS-CoV-2 быстро распространился практически во всех странах. 11 марта ВОЗ официально признала COVID-19 как пандемию. Недавние исследования показали, что SARS-CoV-2 имеет сходную геномную организацию с другими бета-коронавирусами, состоящую из 5'-нетранслируемой области (UTR), репликазного комплекса (orf1ab), кодирующего неструктурные белки (nsps), белок шипика (S) ген, белок оболочки вируса (E), ген мембранного белка (M), ген нуклеокапсидного белка (N), 3'-UTR и несколько неидентифицированных открытых рамок считывания. Хотя SARS-CoV-2 относится к группе бета-коронавирусов, он отличается от MERS-CoV, BCoV и SARS-CoV. Недавние исследования показали, что геном SARS-CoV-2 имеют идентичность нуклеотидов <80 % и сходство нуклеотидов 89,10 % с геномом SARS-CoV [1, 2]. Ещё одним вирусом достаточно близким к возбудителям острого расприраторного синдрома у человека является бычий короновирус (BCoV).

Среди структурных белков CoV S-гликопротеин в большом количестве синтезируется в клетках зараженного животного. Этот белок в инфицированных клетках имеет несколько функций. Он необходим при проникновении вируса в клетку и патогенезе. Субъединица S1 опосредует с помощью его рецептор-связывающий домена взаимодействие вируса с рецептором на поверхности клетки-мишени. S2

субъединица обеспечивает слияние вирусной и клеточной мембран. Также предполагается, что S-белки содержат в себе эпитопы, которые индуцируют синтез иммунной системой животного нейтрализующих антител (Рис. 1.).

Шипики коронавируса существуют в двух разных конформациях: структуры пред слиянием мембраны вируса и клетки-хозяина присутствуют на зрелых вирионах и имеют форму гвоздики с тремя головками, образованными S1-субъеденицами, сидящими сверху на тримерном стебле, состоящем из S2-белка; структуры, образующиеся после слияния, имеют форму гантели, образованной специальной структурой в виде шести спирального пучка из трех субъединиц S2 [2]. Структуры S2 разных родов коронавирусов весьма сходны друг с другом как до, так и после слияния. Достаточно консервативны и их последовательности, что может оказаться полезным для выбора группоспецифичного маркера как белкового, так и нуклеинового. В противоположность S2, субъединицы S1 вирусов, представителей разных роды имеют различную структуру и последовательность. Они предназначены для распознавания различающихся рецепторов клеток различных животных-хозяев того или иного штамма коронавирусов [2]. S1-белок имеет два домена, N-терминальный домен (S1-NTD) и C-терминальный домен (S1-CTD), либо один из них, либо оба могут функционировать как рецептор-связывающий домен. Домены S1 вирусов разных родов имеют свои структурные особенности, которые являются уникальными для рода. S1-CTD особенно разнообразны, и могут обладать очень низким или даже отсутствием структурного сходства у представителей разных родов [2].

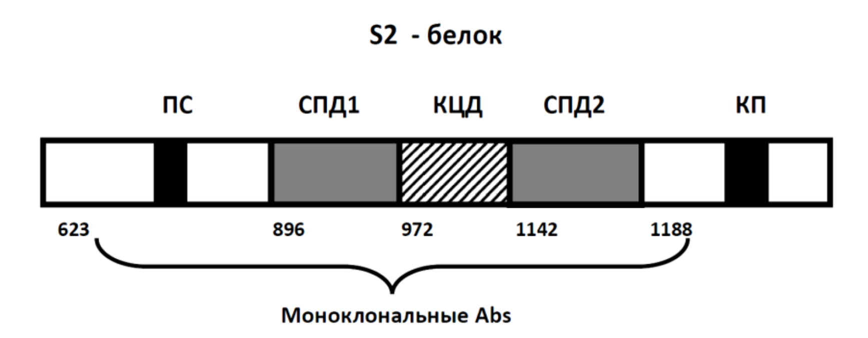


Рисунок 1 - SARS-CoV S2 белок и эпитопы mAb. Отмечены различные функциональные домены SARS-CoV S2 – белка. ПС -пептид слияния, СПД1 и СПД2 – спиральные повторяющиеся домены, участвующие в слиянии, КП – концевой повтор, КЦД – консервативный центральный домен, взятый для исследования в данной работе. Под схемой – координаты в аминокислотах от начала незрелого транслированного продуктиа гена S SARS-CoV. Фигурной скобкой отмечены области, с которыми реагируют моноклональные Abs

Методика исследований. Для поиска гомологов белка S2 бетакоронавирусов использовались три маркерных белка SARS-CoV-2, возбудителя COVID-19, SARS-CoV возбудителя атипичной пневмонии в Китае в 2002 – 2003 годах и BCoV – бычьего коронавируса, близкого к двум предыдущим вирусам. Для сравнения а.к. последовательностей с базами данных NCBI US использовались алгоритмы BLASTp, DELTA-BLAST, PHI-BLAST и PSI-BLAST [3]. Хиты, найденные с использованием алгоритма PSI-BLAST после третей итерации с уровнем статистической достоверности E-value < 0,002, были взяты для дальнейшего изучения. В качестве контроля по аналогичной схеме провели сравнение маркерных белков с последовательностями из хвостатых бактериофагов, бактерий и человека. Был получен файл с а.к. после-

довательностями гомологов S2 бетакоронавирусов в формате FASTA, в котором объединили все найденные последовательности и использовали их для обработки в пакете программ MEGA6 [5]. Выравнивание осуществляли при помощи алгоритма MUSCLE [4]. В анализе приняли участие 77 аминокислотных последовательностей. Эволюционный анализ также был проведен в MEGA6 [5].

Результаты исследований и их обсуждение. Дерево дало весьма любопытные результаты. Среди найденных нами гомологов S-гликопротеина бычьего коронавируса помимо гомологов белков шипиков коронавирусов животных и человека, оказались белки фагов, ответственные за изменение длины их хвоста, последовательности из метагеномов осадочной и планктонной микробиоты океана, нуклеопорины и HSF-bP1 человека, последовательности из бактерий, в том числе и цианобактерий. Можно заметить, что произошло разделение на две крупные ветви. Первая включает в себя практически все найденные последовательности из метагенома, последовательности из фагов, человеческие нуклеопорины. Во вторую вошли S-гликопротеины коронавирусов животных различных видов и человека, а также гомологичные им последовательности, найденные в некоторых бактериях, в том числе и цианобактериях, некоторые последовательности из фагов, а также HSF-bP1 человека. Рассмотрим каждую ветвь подробнее.

В первой ветви также наблюдается разбиение на несколько подветвей. В первую вошли все найденные белки, контролирующие длину хвоста фагов листерий, лактобацилл и стрептококков – TMP (tail tape measure protein). Хоть и данная подветвь имеет достаточно отдаленное расположения от S2-гликопротеинов коронавирусов животных, но сам факт ее наличия может свидетельствовать об их общих корнях происхождения. Также это предположение можно подкрепить общей способностью исследуемого белка и TMP к тримеризации. К TMP наиболее близкой оказалась большая часть последовательностей из океанического метагенома – GOS 4407730, GOS 7848850, GOS 2780498, GOS 965606, GOS 1732414, GOS 1732987, GOS 9424319, METZ01 LOCUS275133 MM и последовательности из осадочного метагенома, образцы которого были взяты рядом с гидротермальным источником замок Локи в Атлантическом океане - LCGC14 2750990 MSM, LCGC14 0719320 MSM. Связь между гомологами среди последовательностей из планктонического микробиома и фаговыми последовательностями объяснить наличием большего количества последних в пикопланктоне. Это могло произойти, в том числе, и в результате попадания сточных вод отходов животноводства в океан, т.к. хостами попавших в эту ветвь фагов являются бактерии, жизнедеятельность которых связана с сельским хозяйством. А вот близкая связь TMP с последовательностями из пелагического микробиома более интересна, т. к. ее наличие может свидетельствовать о более раннем происхождении исследуемого объекта и его гомологов, а также о возможной связи не только с прокариотами и эукариотами, но и с археями.

Вторая подветвь, достаточно близка к первой, состоит из сайт-специфичных рекомбиназ фагов *Erysipelothrix sp.,* белков капсида и белков хвоста некоторых фагов, пары последовательностей из морских метагеномов – GAG72381.1:78-159 MSM, GOS 2883076 MM. Также сюда попали пептидазы и последовательности из бактерий *Desulfovibrionacea*, *Gammaproteobacteria, Laribacter hongkongensis,* также предполагаемая эндопептидаза и муреин-гидролаза фага Mediterranean phage uvMED. Связь последней и в принципе, связь между пептидазами и S2-гликопротеином любопытна, т.к. S2-белок обеспечивает слияние вирусной и клеточной мембраны, а муреин-гидролаза в свою очередь разрушает клеточную стенку бактерий, фаг с ее помощью может производить инфицирование бактерии-

хозяина. В целом, подветвь получалась немного разнородной, это можно объяснить наличием в ней последовательностей из метагеномов.

Третью, пожалуй, самую однородную и интересную подветвь образовали гомологи из нуклеопоринов человека и их изоформы, а также внутри нее есть еще одна маленькая подветвь. Она образована последовательностями из пелагического и морского метагеномов – LCGC14 0381710MSM и GOS 21363, и гипотетическим протеином Arthrobacter phage Colucci. Связующим звеном между данными океаническими последовательности с нуклеопоринами человека судя по дереву является последовательность из *Enterococcus faecium,* предположительно несущую функцию нуклеопорина, а также через последовательность GOS1669449*. Enterococcus faecium* является представителем микрофлоры человека и некоторых млекопитающих, ее связь с морскими последовательностями могла появиться из-за сточных вод. Само наличие человеческих белков-нуклеопоринов в получившемся дереве можно объяснить тем, что S2-гликопротеин несет функции нуклеопорина, помимо этого обеспечивает слияние мембран клетки и вируса. Также остается открытым вопрос, почему в эту подветвь попала последовательность именно из пелагической микробиоты гидротермального источника замка Локи.

Перейдем ко второй ветви. Также, как и в случае с первой, мы здесь видимо разделение на более мелкие подветви. Пожалуй, самая любопытная и сложнообъяснимая из всех ранее описанных связей, состоит из HSF-bP1 человека, последовательностей из цианобактерий *Synechococcus sp.,* из морского метагенома GOS 6072953 MM, GOS 9605623 и из *Streptococcus agalactiae H36B* и Clostridium phage PhiS63. Если связь между последовательностями из метагенома, цианобактериями и фагами вполне объяснима и логична, то близкая связь между *Synechococcus sp* и HSF-bP1 заставляет теряться в догадках.

Эволюционная история была выведена с использованием метода UPGMA [4]. Консенсусное дерево было выведено путем 1000 повторов теста [7] взяты для представления эволюционной истории таксонов проанализировано [7]. Ветви, соответствующие разветвлениям, были воспроизведены только те, которые имелись не менее чем в 50 % повторах теста начальной загрузки. Эволюционные расстояния были вычислены с использованием матричного метода JTT [6] и в единицах числа аминокислотных замен на сайт. В анализе были использованы 77 аминокислотных последовательностей. Все позиции, содержащие пробелы и пропущенные данные, были исключены. Эволюционные анализы были проведены в пакете программ MEGA6 [4].

Аминокислотные последовательности, транслированные из контигов морских метагеномов планктонической микробиоты отмечены белыми кружками, белки микробиоты океанических отложений – черными квадратами, последовательности S2 – белков Coronaviridae отмечены белыми ромбами, последовательности белков бактериофагов черными ромбами, белки бактерий черными треугольниками, нуклеопорины человека – белыми квадратами. MM – обозначены последовательности из океанического планктонического микробиома и MSM – из глубоководного осадочного микробиома.

Само появление HSF-bP1 в качестве гомолога можно объяснить его схожестью с нуклеопоринами человека и на дереве видна эта связь, но почему при этом HSF-bP1 оказалась более близкой последовательностью к исследуемым S2-гликопротеинам, чем нуклеопорины, при условии наличия известных схожих функций между последними – трудно объяснить. Наиболее близкая к ним подветвь, состоящая из пары последовательностей метагенома планктонной микробиоты океана – GOS 9479853 и GOS 3388849.

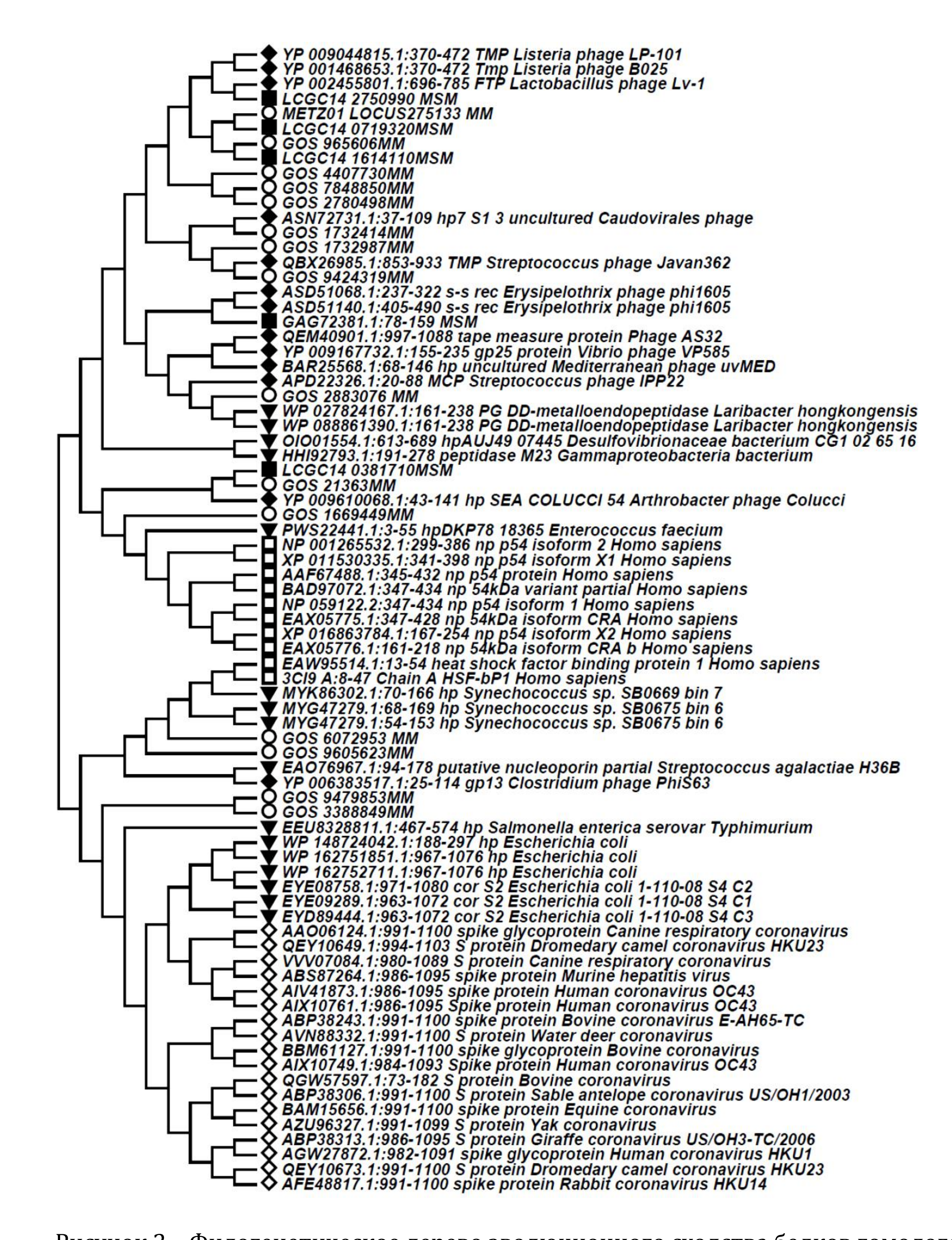


Рисунок 2 – Филогенетическое дерево эволюционного сходства белков гомологов S-гликопротенина бычьего коронавируса из метагеномов микробиоты океана

Она находится посередине ветви и будто связывает ранее описанную подветвь и нижележащую, куда вошли все гомологичные гликопротеиды коронавирусов разных видов животных и их гомологи, найденные в *Escherichia coli* и *Salmonella enterica serovar Typhimurium.* Гомологи последних ближе к метагено-

мам, это скорее всего связано со сточными водами, попадающими в океанические воды. Подветвь с гомологами других коронавирусов достаточно однородна. Стоит отметить, что S-гликопротеины бычьего коронавируса достаточно тесно связаны с коронавирусами человека, в частности самые близкая связь оказалась между белками шипиков Human coronavirus OC43 и Bovine coronavirus. Также шипики бычьего коронавируса достаточно близки с гликопротеидами водяных оленей и антилоп. Белки из еще одного коронавируса человека – Human coronavirus HKU1 имеют самую близкую связь с шипиками коронавируса жирафа.

В целом, можно сказать, что дерево дает понять, что S-гликопротеины коронавирусов образуют отдельную подветвь и достаточно близки между собой. Помимо этого, связь близка между ними и найденными гомологичными последовательностями из бактерий жкт и возбудителями кишечных инфекций, а также с парой последовательностей из метагенома и цианобактерий. Возможно, это связано с попаданием сточных вод от животноводческих хозяйств в океаны, так и наоборот. Отдельную ветвь на дереве образовала большая часть последовательностей из метагенома и белков, контролирующих длину хвоста фагов, интересно, что гомологи из нуклеопоринов оказались наиболее близки именно к ним. Также, данное дерево дало понять, что шипики бычьего коронавируса имеют больше всего сходств с шипиками коронавируса человека. Стоит продолжать филогенетические исследования в данном направлении, чтобы получить однозначные выводы.

Выводы. 1. Проведен поиск океанических гомологов белка S2 шипиков бычьего коронавируса в метагеномах планктонической и осадочной микробиоты мирового океана. Было найдено 15 и 4 гомологов этого вирусного белка, соответственно.

2. В ходе исследования было получено филогенетическое дерево, состоящее из гомологов S2-белка коронавирусов животных и человека, а также гомологов этого белка, а именно, последовательностей из метагеномов осадочной и планктонной микробиоты океана, белков фагов, контролирующих изменение длины их хвоста, нуклеопоринов и HSF-bP1 человека, последовательностей из бактерий, в том числе и цианобактерий.

3. Последовательности из метагенома планктонной микробиоты океана – GOS 9479853 и GOS 3388849 образовали ветвь наиболее близкую S2 бета-коронавирусов и их гомологам, найденным у Escherichia coli и Salmonella enterica serovar Typhimurium

4. Последовательности белков из океанического метагенома – GOS 4407730, GOS 7848850, GOS 2780498, GOS 965606, GOS 1732414, GOS 1732987, GOS 9424319, METZ01 LOCUS275133 MM и последовательности из осадочного метагенома – LCGC14 2750990 MSM, LCGC14 0719320 MSM оказалась наиболее близкими к белку контролирующему длину хвоста у бактериофагов отряда Caudovirales – гомологов S2 – белка бычьего короновируса.

5. Отдельную ветвь на дереве образовала большая часть последовательностей из метагенома и белков, контролирующих длину хвоста фагов. Выяснившаяся связь между последними и S2-белками весьма важна, т.к. данные белки имеют схожие структурные особенности, что дает возможность выдвинуть предположение о взаимосвязи их происхождения и функции.

Исследование было частично поддержано грантом РФФИ №20-54-53018 ГФЕН_а для Зимина А.А. и выполнено им в рамках этого проекта.

Список литературы

1. Доклад ВОЗ. WHO Coronavirus disease (COVID-2019) R&D (https://www.who.int/blueprint/priority-diseases /key-action/novel-co-ronavirus/en Дата обращения 22 апреля 2020 года).

2. S2 SARS rep UniProt (https://www.uniprot.org/uniprot/P0C6X7 – SARS rep UniProt, P0C6X7. Дата обращения 17 апреля 2020).

3. Altschul S.F. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs / S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman // Nucleic Acids Res. 1997 – Sep 1;25(17):3389-402.

4. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792-7.

5. Tamura K. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 / K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar // Mol. Biol. Evol. 2013. V. 30. P. 2725-2729.

6. Jones D.T., Taylor W.R., and Thornton J.M. The rapid generation of mutation data matrices from protein sequences / Jones D.T., Taylor W.R., and Thornton J.M. // Computer Applications in the Biosciences. 1992. 8: 275-282.

7. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap / J. Felsenstein // Evolution. 1985. 39:783-791.

.

DOI
УДК 574.24:591

СПЕЦИФИЧНАЯ КЛАСТЕРИЗАЦИЯ ОКЕАНИЧЕСКИХ БЕЛКОВ DENV С БЕЛКАМИ БАКТЕРИЙ НАЗЕМНЫХ ЖИВОТНЫХ

Карманова Александра Николаевна[1,2]
Зимин Андрей Антонович[1], канд. биол. наук
[1]*Институт биохимии и физиологии микроорганизмов им. Г.К. Скрябина РАН – обособленное подразделение ФИЦ «Пущинский научный центр биологических исследований РАН», г. Пущино, Российская Федерация*
[2]*Вятский государственный университет, г. Киров, Российская Федерация*

В данной работе исследовано эволюционное сходство гомологов DenV из микробиоты пелагики океана и микробиоты наземных животных, в том числе и той, которая является обычным компонентом сточных вод животноводческих предприятий. Филогения данного набора аминокислотных последовательностей гомологов DenV позволяет сделать предположения о различиях и сходствах в резистентности к УФ-излучению у бактерий этих двух экологических ниш. Это открывает фундаментальные возможности прогнозирования в области очистки сточных вод ферм УФ-светом и возможности самоочистки морских вод за счет природной инсоляции от бактериальной составляющей сельскохозяйственной контаминации.

Ключевые слова: устойчивость к ультрафиолету; DenV; очистка сточных вод ферм; микробиота пелагики океана; микробиота наземных животных

SPECIFIC CLUSTERING OF THE OCEAN DENV PROTEINS WITH PROTEINS OF TERRESTRIAL ANIMAL BACTERIA

Karmanova Aleksandra Nikolaevna[1,2]